

Section 7.5N and 7.5G

1

Normal Model as an Approximation to the Binomial Model and Geometric Random Variables

Learning Objectives –

Normal Model as an Approximation to the Binomial Model

After this section, you should be able to...

DETERMINE whether the conditions have been met to use to use the Normal Model as an Approximation to the Binomial Model.

THEN:

- ✓ COMPUTE and INTERPRET probabilities in context
- ✓ CALCULATE the mean and standard deviation and INTERPRET these values in context

Normal Approximation for Binomial Distributions

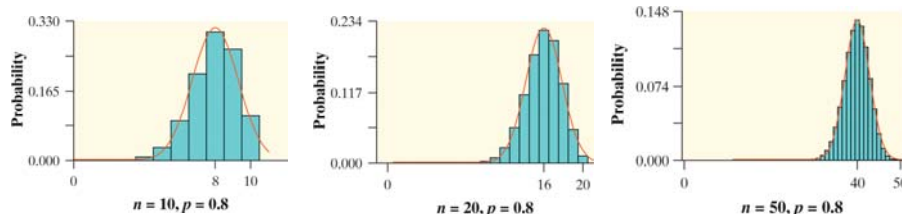
2

As n gets larger, something interesting happens to the shape of a binomial distribution.

The figures below show histograms of binomial distributions when $p=.8$ and different values of n .

- 1) $n=10$
- 2) $n=20$
- 3) $n=50$

What do you notice as n gets larger?



Binomial Random Variables

+ ■ Normal Approximation for Binomial Distributions

3

Binomial Random Variables

Normal Approximation for Binomial Distributions

Suppose that X has the binomial distribution with n trials and success probability p .

When n is large, the distribution of X is approximately Normal with mean and standard deviation

$$\mu_X = np \qquad \sigma_X = \sqrt{np(1-p)}$$

We can use the Normal approximation when:

- $np \geq 10$ and
- $n(1-p) \geq 10$.
- That is, the expected number of successes and failures are both at least 10.

Note,
the binomial distribution would be **MORE accurate**
than the normal distribution.

+ ■ Attitudes toward Shopping

4

Binomial Random Variables

EXAMPLE – Normal Approximation for Binomial Distributions

Sample surveys show that fewer people enjoy shopping than in the past. A survey asked a nationwide random sample of 2,500 adults if they agreed that "I like buying new clothes, but shopping is often frustrating and time-consuming." 60% of the adults surveyed agreed with this statement. What is the probability that between 1,000 and 2,000 of the sample agrees?

Use the following steps to solve this problem:

Steps:

- Check the BINS conditions
- Define the Random Variable
- Check the conditions: $np \geq 10$ and $n(1-p) \geq 10$ ** IMPORTANT: You MUST show both calculations to indicate you verified this condition !!!!
- Calculate the mean and the standard deviation.
- Draw a picture. Then calculate the probability of interest
- State your conclusion, in the context of the problem.



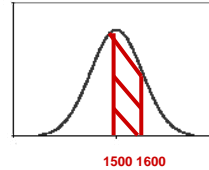
Attitudes toward Shopping

EXAMPLE – Normal Approximation for Binomial Distributions

Solution:

Steps:

- Check the BINS conditions. They check: there are 2 outcomes (agree/disagree), fixed number ($n=2500$), fixed probability ($=.6$), and independent (random sample).
- Define the Random Variable: X =the number in the sample that agree
- Check the conditions: $np \geq 10$ and $n(1-p) \geq 10$ ** IMPORTANT: You MUST show both calculations to indicate you verified this condition !!!!
 $np = 2500(.6) = 1500 \geq 10 \checkmark$
 $n(1-p) = 2500(.4) = 1000 \geq 10 \checkmark$
- Calculate the mean and the standard deviation.
 $\mu = np = 2500(.6) = 1500$
 $\sigma^2 = npq = 2500(.6)(.4) = 600 \quad \sigma = 24.494$
- Draw a picture. Then calculate the probability of interest
 $P(1500 \leq X \leq 1600) = \text{normalcdf}(1500, 1600, 1500, 24.49) = .499$
- State your conclusion, in the context of the problem.
There is about a 50% chance that between 1,500 and 1,600 of the survey respondents agree with the shopping survey question.



5

Binomial Random Variables



6

+ Section 7.5G Geometric Random Variables

7

Geometric Random Variables

Learning Objectives

After this section, you should be able to...

- ✓ DETERMINE whether the conditions for a geometric setting are met
- ✓ CALCULATE probabilities involving geometric random variables

■ Geometric Settings

8

+

Geometric Random Variables

In a binomial setting, the number of trials n is fixed and the binomial random variable X counts the number of successes. In other situations, the goal is to repeat a chance behavior *until a success occurs*. These situations are called **geometric settings**.

Definition:

A **geometric setting** arises when we perform independent trials of the same chance process and record the number of trials until a particular outcome occurs. The four conditions for a geometric setting are

- B** • **Binary?** The possible outcomes of each trial can be classified as “success” or “failure.”
- I** • **Independent?** Trials must be independent; that is, knowing the result of one trial must not have any effect on the result of any other trial.
- T** • **Trials?** **The goal is to count the number of trials until the FIRST success occurs.**
- S** • **Success?** On each trial, the probability p of success must be the same.

■ Geometric Random Variable

In a geometric setting, if we define the random variable Y to be the number of trials needed to get the first success, then Y is called a **geometric random variable**. The probability distribution of Y is called a **geometric distribution**.

Definition:

The number of trials Y that it takes to get a success in a geometric setting is a **geometric random variable**. The probability distribution of Y is a **geometric distribution** with **parameter p** , the **probability of a success on any trial**. The possible values of Y are 1, 2, 3,

Note: Like binomial random variables, it is important to be able to distinguish situations in which the geometric distribution does and doesn't apply!

The Birthday Game

EXAMPLE – Geometric Distributions

Your teacher is planning to give you 20 problems for homework. She agrees to play the Birth Day Game. She secretly thinks of a person's birthday and writes down the day of the week they were born hidden from the students. She then selects at random a student from the class and asks them to guess the day of the week. If the student guesses correctly, the students have only 1 HW problem. If the student guesses incorrectly she repeats the process (selects a different friend and randomly selects any student in the class). The teacher assigns the number of problems to the number of students that guessed.

1) Define the random variable:

Y = the number of guesses it takes to correctly identify the birth day of one of your teacher's friends.

2) Verify that Y is a geometric random variable.

B: Success = correct guess, Failure = incorrect guess

I: The result of one student's guess has no effect on the result of any other guess.

N: We're counting the number of guesses up to and including the first correct guess.

S: On each trial, the probability of a correct guess is $1/7$.



3) What is the probability the student guesses correctly

➤ The 1st student? $P(Y=1) = \underline{\hspace{2cm}} = \underline{\hspace{2cm}}$

➤ The Second? $P(Y=2) = \underline{\hspace{2cm}} = \underline{\hspace{2cm}}$

➤ The Third? $P(Y=3) = \underline{\hspace{2cm}} = \underline{\hspace{2cm}}$

Notice the pattern?

The Birthday Game

11

EXAMPLE – Geometric Distributions



(3 cont.) Probability the student guesses correctly

$$P(Y = 1) = 1/7$$

$$P(Y = 2) = (6/7)(1/7) = 0.1224$$

$$P(Y = 3) = (6/7)(6/7)(1/7) = 0.1050$$

The Pattern... What is the probability the k^{th} student guesses correctly?

Geometric Probability – G(p)

If Y has the geometric distribution with probability p of success on each trial, the possible values of Y are 1, 2, 3, If k is any one of these values,

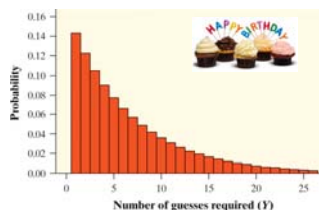
$$P(Y = k) = (1 - p)^{k-1} p$$

The Birthday Game

12

EXAMPLE – MEAN of the Geometric Distributions

- The table and histogram below shows part of the probability distribution of Y .
- We can't show the entire distribution because the number of trials it takes to get the first success could be an incredibly large number.



y_i	1	2	3	4	5	6	...
p_i	0.143	0.122	0.105	0.090	0.077	0.066	

Shape: The **heavily right-skewed shape** is characteristic of any geometric distribution. That's because the most likely value is 1.

Center: We could calculate the expected value $E(Y) = \sum y_i p_i$ and the mean is $\mu_Y = 7$.
We'd expect it to take 7 guesses to get our first success.

Spread: The standard deviation of Y is $\sigma_Y = 6.48$. **If the class played the Birth Day game many times, the number of homework problems the students receive would differ from 7 by an average of 6.48.**

Mean (Expected Value) of Geometric Random Variable

If Y is a geometric random variable with probability p of success on each trial, then its **mean** (expected value) is $E(Y) = \mu_Y = 1/p$.

Geometric Random Variables

+ Summary of Geometric Distributions

1) A geometric setting consists of

- ✓ repeated trials of the same chance process in which each trial results in a success or a failure;
- ✓ trials are independent;
- ✓ each trial has the same probability p of success;
- ✓ and the goal is to count the **number of trials until the first success occurs.**

2) Define the geometric random variable:

- ✓ Y = the number of trials required to obtain the first success. Y are the positive integers 1, 2, 3,

3) The Geometric probability distribution model: $G(p)$

- ✓ p = probability of success
- ✓ $q = 1 - p$ = probability of failure
- ✓ Y = number of trials until the first success occurs

$$P(Y = k) = (1 - p)^{k-1} p \quad E(X) = \mu = \frac{1}{p} \quad \sigma = \sqrt{\frac{q}{p^2}}$$

+ Section 7.5: The 3 Types of Distributions

Summary

- ✓ **Binomial model**
 - When we're interested in the number of successes in a fixed number of trials, with a fix probability, and 2 outcomes (binomial).
- ✓ **Normal model**
 - To approximate a Binomial model when we expect at least 10 successes and 10 failures.
- ✓ **Geometric model**
 - When we're interested in the number of Binomial trials until the **FIRST success.**

+ Binomial Distributions in Statistical Sampling

Learning Objectives

After this section, you should be able to...

- ✓ DETERMINE whether the conditions are met using a binomial distribution
- ✓ CALCULATE probabilities of success using a binomial distribution

Sampling Without Replacement Condition

When taking an SRS of size n from a population of size N , we can use a binomial distribution to model the count of successes in the sample as long as

$$n \leq \frac{1}{10} N$$

+ Binomial Distributions in Statistical Sampling

The binomial distributions are important in statistics when we want to make inferences about the proportion p of successes in a population.

Suppose 10% of CDs have defective copy-protection schemes that can harm computers. A music distributor inspects an SRS of 10 CDs from a shipment of 10,000.

Let X = number of defective CDs.

What is $P(X = 0)$? *Note, this is not quite a binomial setting. Why?*

The actual probability is $P(\text{no defective}) = \frac{9000}{10000} \cdot \frac{8999}{9999} \cdot \frac{8998}{9998} \cdots \frac{8991}{9991} = 0.3485$

Using the binomial distribution, $P(X = 0) = \binom{10}{0} (0.10)^0 (0.90)^{10} = 0.3487$

In practice, the binomial distribution gives a good approximation as long as we don't sample more than 10% of the population.